



Capturing Deformations of Interacting Non-rigid Objects Using RGB-D Data

Antoine Petit, Stéphane Cotin, Vincenzo Lippiello, Bruno Siciliano

► To cite this version:

Antoine Petit, Stéphane Cotin, Vincenzo Lippiello, Bruno Siciliano. Capturing Deformations of Interacting Non-rigid Objects Using RGB-D Data. IROS 2018 - IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct 2018, Madrid, Spain. hal-01889771

HAL Id: hal-01889771

<https://inria.hal.science/hal-01889771>

Submitted on 8 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Capturing Deformations of Interacting Non-rigid Objects Using RGB-D Data

Antoine Petit¹, Stéphane Cotin¹, Vincenzo Lippiello², Bruno Siciliano²

Abstract—This paper presents a method for tracking multiple interacting deformable objects undergoing rigid motions, elastic deformations and contacts, using image and point cloud data provided by an RGB-D sensor. A joint registration framework is proposed, based on physical Finite Element Method (FEM) elastic and interaction models. It first relies on a visual segmentation of the considered objects in the RGB images. The different segmented point clouds are then processed to estimate rigid transformations with an ICP algorithm, and to determine geometrical point-to-point correspondences with the meshes. External forces resulting from these correspondences and between the current and the rigidly transformed mesh can then be derived. It provides both non-rigid and rigid data cues. A classical collision detection and response model is also integrated, giving contact forces between the objects. The deformations of the objects are estimated by solving a dynamic system balancing these external and contact forces with the internal or regularization forces computed through the FEM elastic model. This approach has been here tested on different scenarios involving two or three interacting deformable objects of various shapes, with promising results.

I. INTRODUCTION

The problem of capturing the deformations of non-rigid objects is an active research field in computer vision that finds prospective applications in robotic manipulation, human/machine and tangible interactions, or motion capture. It becomes a challenging issue when considering interacting multiple objects. Accurately and reliably registering the interactions can be crucial in applications such as bin picking tasks of soft objects, grasping non-rigid objects with a soft manipulator, to handle food for instance, or performing robotic surgical operations in complex deformable environments with interacting organs or tissues.

Although the case of rigid or articulated objects in close interaction has shown an increasing interest in the research community in the recent years, considering deformable objects remains an open problem. The parameter space becoming much larger than 3D rigid poses or articulation joint states, several challenges must be overcome to jointly estimate the deformations of the interacting non-rigid objects. A main one regards inter occlusions between objects and deformations due to contacts. Tracking independently the different objects would indeed result, due to inaccuracies

and ambiguities in the visual or point cloud information, in inter-penetrations and physically unrealistic behaviours between objects. Instead, a simultaneous registration method relying on modelling the contacts between the objects would inherently handle these issues and reach a reliable behavior.

Following this idea, we introduce a solution to integrate collision detection and response models into a rigid and non-rigid registration framework based on successive RGB-D data. This work extends the approach proposed in [26, 25] which is designed for single objects with FEM elastic deformation model. Collisions are handled by hierarchical bounding volume to contact point detection methods and on penalty forces to compute the contact response.

The method described here has been tested on scenarios involving two or three deformable objects, the scenes being observed by commodity RGB-D sensors.

The paper is organized as follows: after a review of the related literature, along with the motivations for the proposed system in section II, descriptions of the elastic and interaction models are provided in section III and IV. Section V introduces the multiple objects registration framework and some experimental results are reported in section VI to illustrate and validate the approach.

II. RELATED WORKS AND MOTIVATIONS

The goal is here to continuously estimate, based on successive RGB-D frames, the deformations undergone by multiple objects, given their known meshes. In the literature, various approaches have been proposed to register deformable objects, using vision and/or range data, and they can be classified according to the underlying model of the considered object, the resolution method, and according to the extracted visual and geometric features.

A. Single object non-rigid registration

Based on implicit physical models, approaches in [17, 3, 28] use a 1D parametric curve or 2D splines models (B-splines, Radial Basis Functions) to track single deformable objects in monocular images. This class of methods relies on the minimization of an energy function involving an external energy term related to some image features, and an internal energy term regularizing curvature, bending or twisting, compelling the model to vary smoothly. Adapting these techniques to register 3D shapes or surfaces in monocular images is much complex, since 3D deformations can imply ambiguous 2D transformations, resulting in an underconstrained problem. A first attempt by Terzopoulos *et al.* [34], relying on 3D splines and inspired by Kass *et*

¹A. Petit and S. Cotin are with Inria, Mimesis Group, 1 place de l'hôpital, 67000 Strasbourg, France.

²V. Lippiello and B. Siciliano are with Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione, Università degli Studi di Napoli Federico II, via Claudio 21, 80125 Napoli, Italy, email: {antoine.a.petit,stephane.cotin}@inria.fr, {vincenzo.lippiello,bruno.siciliano}@unina.it. This research has been partially funded by the EC Seventh Framework Programme (FP7) within RoDyMan project 320992.

al. [17], densely processes gradient features, to compute the data energy term. Less ambiguous feature-based approaches [29] have been preferred and additional constraints are often added to solve ambiguities. With point cloud data, methods in [15, 36] employ an RGB-D sensor to register the acquired point cloud to a surface mesh by minimizing an error function accounting for geometric or direct depth and color errors, and a stretching penalty function for the mesh. By means of a NURBS parametrization [15] or an optimized GPU implementation [36], real-time performance can be achieved. Although these two systems have shown promising and impressive results, they are still limited to isometric or small elastic deformations, by means of regularization functions proportional to squared distances between nodes of the mesh. Here, we wish to model elasticity in a more physically realistic manner, in order to handle volumetric effects and larger strains.

Indeed, another formulation of the problem relies on physics-based deformable models to perform registration. Other sorts (such as non-linear elasticity) and magnitudes of deformations can then be handled, inferring more consistently shape and/or volumetric regularization. The solution can be determined by balancing internal and external forces or, equivalently, minimizing energy functions. Physics-based methods include discrete mass-spring-damper systems [7, 31], or more explicit approaches relying on the Finite Element Method (FEM), based on continuum mechanics. In [31], based on a mass-spring-damper system, 3D-3D correspondences, determined through a probabilistic inference, enable the computation of the external forces applied to the mesh. However mass-spring systems are limited in terms of magnitude of elastic deformations, and do not guarantee certain mechanical properties such as volume preservation, failing to handle volumetric effects. First attempts for registration employing the FEM for 3D surfaces in [5, 21] used linear models. More recently in [20], registration in monocular images is addressed by designing a stretching/shrinking energy using continuous mechanical constraints on 2D elements assuming linear elasticity, and some 3D boundary conditions. Haouchine *et al.* [13] use a linear tetrahedral co-rotational FEM model, coping with larger elastic deformations, external forces being related to correspondences between tracked 3D feature points mapped to the 3D mesh by means of a stereo camera system. This system has been then extended in [12], involving a non-linear St-Venant Kirchhoff FEM model to efficiently address hyperelastic behavior, registration being addressed through monocular keypoint based external forces.

B. Multiple objects registration

Estimating simultaneously the 3D transformations undergone by multiple objects with vision data is already a well studied problem for rigid or articulating objects. Consisting in retrieving the 6D pose of all rigid objects or the various joint angles in the case of articulated objects, common approaches address the task by independently processing the different objects [33, 2]. Other works propose to register

objects and capture interactions, and to cope with inter occlusions. In the case of articulated hand-hand, hand-object and object-object interactions, methods are numerous [23, 11, 24, 27, 19, 32, 18, 30]. In these works the interactions are modelled through physics-based constraints, enabling to properly deal with collisions and occlusions, using single or multiple RGB or RGB-D cameras. However, to our knowledge, handling multiple deformable objects in interaction is a quite left over problem. We can mention the recent system designed in [10], which introduces a collision detection process in its non-rigid reconstruction pipeline, to deal with auto-collisions. However this technique falls in the class of non-rigid surface reconstruction of the whole observed scene. Our method, as a template tracking technique, would instead explicitly consider different known deformable interacting entities, and in a volumetric manner.

C. Contributions

In addition to the generation of occlusions, interactions and contacts also result in deformations on the considered objects. Therefore, we advocate for the use of a physics based approach to model both the deformations and the interactions, by relying on volumetric FEM-based models, as proposed in [26, 25] for single elastic objects, and on collision detection models. Based on successive RGB-D data inputs, the main contribution of this paper is thus the introduction of a collision detection and response model. Another improvement lies in an estimation of rigid transformations into the FEM resolution, by the integration of rigid ICP-based elastic forces.

D. Overview of the system

As illustrated in Figure 1, our frame-by-frame tracking system can be outlined as follows:

Input: the known 3D volumetric meshes of N_O objects, a given RGB-D frame, and assuming correct registrations at the previous frame.

- 1) Visual segmentation of the considered objects, with a parallel graph cut-based approach on each object.
- 2) Using the resulting segmented point clouds, perform rigid Iterative Closest Point (ICP) algorithms to estimate the rigid transformations from the point clouds to the meshes.
- 3) Derive linear external elastic forces between the current meshes and both the rigidly transformed meshes and the segmented point clouds, using closest point correspondences.
- 4) Numerical resolution of mechanical equations to compute the deformations on the objects, based on tetrahedral linear co-rotational FEM models and on collision detection and response models.

III. ELASTIC DEFORMATION MODEL

We assume the considered objects to be elastic and isotropic, and we employ for each one a co-rotational Finite Element Method elastic model, as in [26, 25]. It provides

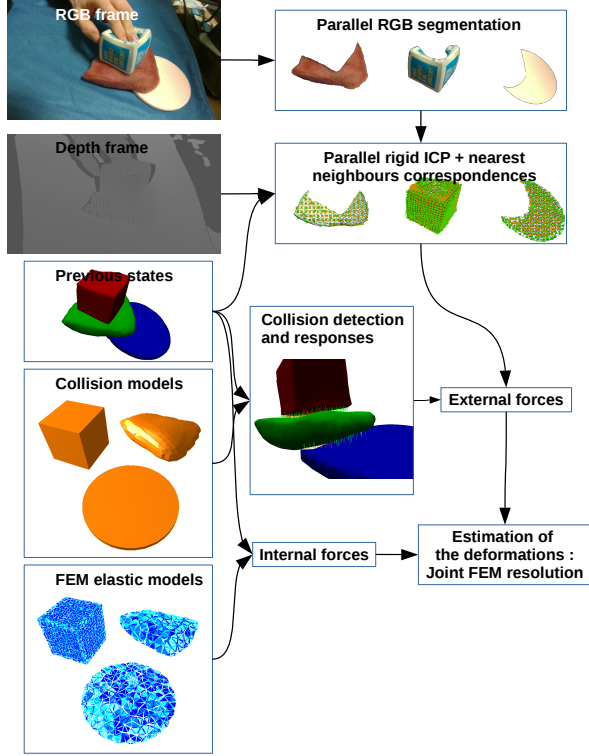


Fig. 1: Proposed pipeline for multiple deformable objects tracking.

physical realism, by relying on continuum mechanics, while being computationally efficient.

Let us note that the modularity of the implemented framework would allow us to use other physical deformation models (hyperelastic, viscous, isotropic or anisotropic, fractures...).

For an exhaustive description of the FEM, the reader can refer to [6]. As a concise description, the method consists in tessellating the deformable object into a mesh made of elements connecting a set $\mathbf{X}_i = \{\mathbf{x}_{i,j}\}_{j=1}^{n_{X_i}}$ of 3D vertices. The deformation fields over the elements are approximated as continuous interpolations of the displacements of the vertices. We rely here on a volumetric linear FEM approach with tetrahedral elements.

For elasticity modelling, we then resort to the linear elasticity theory, with Hooke's law, and to the infinitesimal strain theory [6], modified by adopting a co-rotational approach [8] so as to accommodate to rotation transformations.

If the deformations $\hat{\mathbf{u}}_e$ can be written as $\hat{\mathbf{u}}_e = \mathbf{x}_e - \mathbf{x}_{e,0}$, we define here $\hat{\mathbf{u}}_e^r = \mathbf{R}_e^{-1}\mathbf{x}_e - \mathbf{x}_{e,0}$, with $\mathbf{R}_e^{-1}\mathbf{x}_e$ the back rotated deformed coordinates of the four vertices of e , stacked into the 12×1 vector \mathbf{x}_e . \mathbf{R}_e is a 12×12 block diagonal matrix containing four copies of the 3×3 rotation matrix corresponding the rotational component of the deformations of the element.

The internal elastic forces \mathbf{f}_e exerted on the vertices of e can then be related to $\hat{\mathbf{u}}_e^r$ through:

$$\mathbf{f}_e = \mathbf{R}_e \mathbf{K}_e \hat{\mathbf{u}}_e^r \quad (1)$$

with \mathbf{K}_e the stiffness matrix of the element. \mathbf{K}_e depends on two elastic parameters of the material, the Young modulus E_i and the Poisson ratio ν_i , which are specific to each involved object. With this technique, the overall forces on the whole mesh can be summed to zero, while computational efficiency is ensured since \mathbf{K}_e can be computed in advance, in contrast to non-linear FEM approaches.

IV. INTERACTION MODEL

Modelling the interactions between the considered objects constitutes a central issue to allow a joint registration of the deformations. We propose here a state-of-the-art model consisting in classical collision detection and collision response strategies. Collision being performed on surface meshes, we rigidly map a collision triangular surface mesh of each object on their corresponding volumetric tetrahedral mesh.

A. Collision detection

Collisions can be handled by employing different approaches to determine the intersections and the contacts between the interacting objects. Reviews of collision detection methods can be found in [14, 35]. Intersections are computed by performing successive distance tests, based on a hierarchical set of bounding volumes around the polyhedral triangular collision meshes.

Here, we employ hierarchies of axis aligned bounding boxes (AABB) around the considered objects. First, as a broad phase, intersections between the coarsest AABBs are found. Then, as a narrow phase, we descend in the AABB hierarchies of the objects in a coarse-to-fine manner. Within the finest AABBs, a search for intersections between pairs of geometric primitives is performed. The primitives are triangles in our case. The computation of intersections is based on a threshold on the distances between the primitives (triangle-to-triangle), and returns corresponding pairs of collision (or contact) points on the colliding primitives.

B. Collision response

In order to model the contact forces between two colliding deformable objects i and j , we use a simple penalty-based method. For two detected collision points \mathbf{p} and \mathbf{q} on the collision surfaces of the interacting objects, barycentric mappings on the tetrahedral meshes are carried out to determine two corresponding points $\mathbf{x}_i(\mathbf{p})$ and $\mathbf{x}_j(\mathbf{q})$, in both volumetric models. Then a penetration distance δ can be defined as $\delta = d - D$, with $d = \|\mathbf{x}_i(\mathbf{p}) - \mathbf{x}_j(\mathbf{q})\|$ and D a proximity threshold. As a collision normal $\mathbf{n}_{i,j}$ we choose $\mathbf{n}_{i,j} = \frac{\mathbf{p} - \mathbf{q}}{\|\mathbf{p} - \mathbf{q}\|}$.

A spring-like repulsive penalty force $\mathbf{f}_{j/i}^c(\mathbf{x}_i)$, exerted by \mathbf{x}_j on \mathbf{x}_i , can then be computed as:

$$\mathbf{f}_{j/i}^c(\mathbf{x}_i) = -k\delta\mathbf{n} \quad (2)$$

with k a stiffness factor.

V. MULTIPLE OBJECTS REGISTRATION

The frame-by-frame registration problem that we address consists in fitting the point cloud data, provided by an RGB-D sensor, with the tetrahedral meshes, in terms of both rigid and non-rigid transformations. The method proposed in [26] for single objects is here extended to multiple objects. Consequently, point cloud segmentation, matching, rigid and non-rigid registration techniques are adapted, as described below.

Preliminary parallel visual segmentation.

The visual segmentation step presented in [26], based on a graph cut minimization over foreground/background color models, is first performed. The objective is to restrict the acquired point clouds to the considered objects, so as to avoid ambiguities in the matching process, with the background or with occluding shapes. Dealing with multiple objects, the process is carried out in a parallel manner on each object O_i , giving 3D target point clouds $\{\mathbf{Y}_i\}$. We limit the size of each \mathbf{Y}_i by sampling it on a regular grid in the image plane.

Parallel rigid pose estimation

Each mesh \mathbf{X}_i , initially in its state computed for the previous frame, is then rigidly registered to each observed segmented point cloud \mathbf{Y}_i . In contrast to the approach proposed in [25, 26], for which the rigidly transformed meshes are used to initialize the non-rigid phase, we derive elastic forces attracting towards them. The idea is to integrate these rigid transformations in a smooth and simultaneous manner into the FEM simulation framework, to properly handle collisions. We actually do not directly consider \mathbf{X}_i but its vertices which appears visible from the previous estimated state in the camera frame. We define $\mathbf{X}_{V,i}$ as the visible vertices of \mathbf{X}_i , lying on the surface of the object. $\mathbf{X}_{V,i}$ is determined by a visibility test on the rendered mesh, with respect to its previous state. It is performed on the whole rendered scene to consider occlusions between the interacting object.

As in [25, 26] a classical rigid ICP algorithm [4] is then carried out for each object, between \mathbf{Y}_i and $\mathbf{X}_{V,i}$. We can compute the rigidly transformed mesh $\mathbf{X}_i^t = \{\mathbf{x}_{i,j}^t\}_{j=1}^{n_{\mathbf{x}_i^t}}$ through:

$$\mathbf{x}_{i,j}^t = \mathbf{T}_i \mathbf{x}_{i,j} \quad (3)$$

with \mathbf{T}_i the rigid transformation estimated with the ICP. Elastic forces $\mathbf{f}_{ext,i}^r$ between both sets \mathbf{X}_i^t and \mathbf{X}_i can then be written as:

$$\mathbf{f}_{ext,i}^r(\mathbf{x}_{i,j}) = k_{i,ext}^r(\mathbf{x}_{i,j}^t - \mathbf{x}_{i,j}) \quad (4)$$

with $k_{ext,i}^r$ a stiffness parameter.

Parallel point cloud matching for non-rigid registration

With the aim of fitting the meshes to the point clouds in a non-rigid manner, still based on [26], we compute external forces that are exerted by the observed segmented point clouds \mathbf{Y}_i on the different corresponding sets $\mathbf{X}_{V,i}$. We use sets of external forces $\mathbf{f}_{ext,i}^d$ related to geometrical information, as introduced in [26]. The method consists first in determining nearest neighbors correspondences, both from

the segmented point clouds to the meshes and from the meshes to the segmented point clouds.

For each object, based on the two sets of mesh-to-point cloud and point cloud-to-mesh correspondences, we can compute an external elastic force $\mathbf{f}_{ext,i}^d$ exerted on each $\mathbf{x}_{i,j}$ in $\mathbf{X}_{V,i}$ as follows:

$$\mathbf{f}_{ext,i}^d(\mathbf{x}_{i,j}) = k_{ext,i}^d(\mathbf{x}_{i,j} - \mathbf{y}_{i,j}^f) \quad (5)$$

where $\mathbf{y}_{i,j}^f$ is a linear combination of points in the point clouds which are matched to $\mathbf{x}_{i,j}$, either from mesh-to-point cloud and from point cloud-to-mesh correspondence sets. $k_{ext,i}^d$ is the stiffness of these external elastic forces, which is here common to each object.

Resolution of the deformations

The objective of the resolution process is to balance the sets of external forces capturing rigid, non-rigid transformations of the objects and the interactions between themselves, with the internal forces based on the different deformation FEM models presented in Section III. The estimation consists in solving for each object a dynamic system of linear ordinary differential equations involving the internal and the different external forces, based on Lagrangian dynamics:

$$\mathbf{M}_i \ddot{\mathbf{x}}_i + \mathbf{C}_i \dot{\mathbf{x}}_i + \mathbf{f}_i = \mathbf{f}_{ext,i}^d + \mathbf{f}_{ext,i}^r + \mathbf{f}_i^c \quad (6)$$

with $\mathbf{f}_i = \mathbf{K}_i \mathbf{x}_i + \mathbf{f}_{0,i}$

where \mathbf{x}_i is a $3n_{X_i} \times 1$ vector containing the positions to estimate of the vertices of each object O_i . \mathbf{M}_i and \mathbf{C}_i are the $3n_{X_i} \times 3n_{X_i}$ mass and damping matrices, \mathbf{K}_i the $3n_{X_i} \times 3n_{X_i} N_o$ global stiffness matrix which sums the $3n_{X_i} \times 3n_{X_i}$ element-wise *rotated* stiffness matrices $\mathbf{K}_e^r = \mathbf{R}_e \mathbf{K}_e \mathbf{R}_e^{-1}$, written with respect to whole set of vertices, and $\mathbf{f}_{0,i}$ the corresponding global offset summing the element-wise ones $\mathbf{R}_e \mathbf{K}_e \mathbf{x}_{e,0}$. $\mathbf{f}_{ext,i}^d$, $\mathbf{f}_{ext,i}^r$ and \mathbf{f}_i^c are $3n_{X_i} \times 1$ vectors containing the external forces and resulting from the interaction model and the collision responses between object O_i and the other objects.

An Euler implicit integration scheme is used to solve the system with respect to each \mathbf{x}_i , along with a conjugate gradient method. We could consider the static case, with the static equilibrium of the deformations assumed to be reached, so the transient and the dynamic terms of equation (6) would be neglected, leading to simply solving the equality between internal elastic forces and external forces with the conjugate gradient method.

VI. RESULTS

The results presented here report different scenarios involving objects of different shapes, potentially textureless, and with different sort of interactions.

The first sequence indeed shows a stuffed toy and elastic rubber toy brain colliding, the second a silicon elastic disk wrapping a silicon crocodile puppet filled with plastic balls. In the third sequence, the disk, a silicon object and cube of foam are stacked and squeezed together.

The point clouds of the investigated scene are acquired using different structured light commodity RGB-D cameras,

an Asus Xtion: 320×240 RGB and depth images being processed, for the first sequence, and an Intel RealSense SR300 (960×540 images) for the second and third test sequence.

In order to create the geometric and deformation models of some of the considered objects (the squirrel, the crocodile, the toy brain, the light red silicon object), surface meshes of the undeformed objects were first reconstructed offline. We have used the RGB-D based dense 3D reconstruction techniques KinectFusion [22] and InfiniTAM [16] which is implemented for the RealSense SR300. The large resolution and short range of the RealSense SR300 with respect to the Asus Xtion enables to capture and model finer grain geometries, such as the one of the crocodile. Finally, some remeshing and smoothing tasks have been carried out with a modelling engine, to get fair, closed and clean surface meshes of the objects. The corresponding volumetric tetrahedral meshes have been generated by achieving a 3D Delaunay triangulations within the surface meshes, thanks to the CGAL library². As an approximation, we assume the isotropy of the materials, in order to apply the deformation model described in Sect. III. In Figure 2 are featured some the considered tetrahedral meshes, as well as the visual and collision surface models for the crocodile.

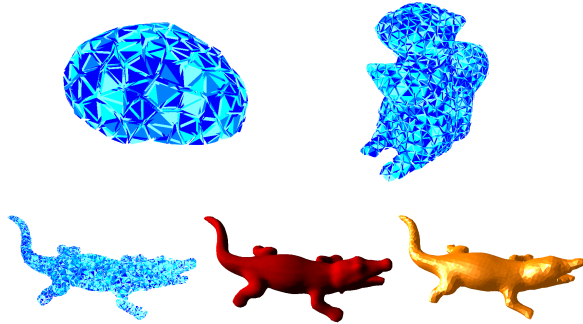


Fig. 2: Tetrahedral meshes for the brain, squirrel and crocodile, plus the visual and collision models for the crocodile toy.

The software implementation relies on the Simulation Open Framework Architecture (SOFA) [9], which provides the FEM elastic and the collision models.

Qualitative results are shown in Figure 3 for the different scenarios, as in the provided video. Let us note that initialization is performed manually, rigidly, close to the actual initial poses and we assume that the objects are not initially in contact.

The case of the silicon disk wrapping the crocodile toy is to be noticed. Indeed, the collision between both objects enables to constraint sufficiently the silicon disk so that it gets registered coherently, even with severe bending deformations, with occlusions by the crocodile, and with errors in the segmentation.

In Figure 4 we can visualize the vertices of the visible surface of the meshes (green squares), which are processed for registration with the segmented point clouds (brown squares). Different samplings for the point clouds are used, accordingly to the resolutions of the meshes. The detected collisions between the interacting objects can be seen on the third row with segments between the found contact points.

The focus on the collision regions in Figure 5 with wireframe views illustrates that contacts and occlusions between the objects can be properly handled, when objects are squeezed, wrapped or stacked together. Independently tracking them leads to inter-penetrations.

In Figure 6 are shown some failure situations when forces resulting from the rigid ICP registration are not integrated in the resolution.

As for computational issues, segmentation being addressed in a parallel manner, and the integration of collision requiring extra computations when object are getting into proximity, the method does not achieved fast performance. $5 - 10fps$ framerates are reached, depending on the number of objects and on the resolutions of the collision surface meshes.

VII. DISCUSSION

The experiments presented above show some promising results for the method. However a few limitations of the method shall be discussed. The system requires some tuning for both the FEM models of the objects and the collision detection and response models to obtain stability and a proper behavior. Some instabilities or ripples can indeed appear when the involved object undergo fast rigid motions or when to many interactions are involved, for instance when two many objects of complex shapes and collision models are involved. In these cases the addition of numerous and potentially large and opposite external forces coming from the point cloud correspondences, the rigid motion estimation and the collisions responses can make the Lagrangian dynamics equations hardly to be solved using the classical implemented Conjugate Gradient solver. Thus in these cases some tuning can be required to achieve a successful resolution of the whole system, and contact detection distances need to be set quite high in some cases (see Figure 4, third row).

The collision detection and response models are also very classical and a more elegant approach, guaranteeing a more physical response, could be undertaken, by resorting to the Signorini's model and by solving the contacts as a Linear Complementary Constraint problem [1]. Including friction through Coulomb's law could also provide some improvements.

Although the system has proven to be fast when considering a single object, with multiple objects computations linearly increase with the number of objects, and even more when contacts occur. One major burden in the computations lies actually in the segmentation process which is handled independently for each considered object. A simultaneous segmentation method, in a pyramidal manner, could be designed to obtain a better efficiency.

²<http://www.cgal.org>

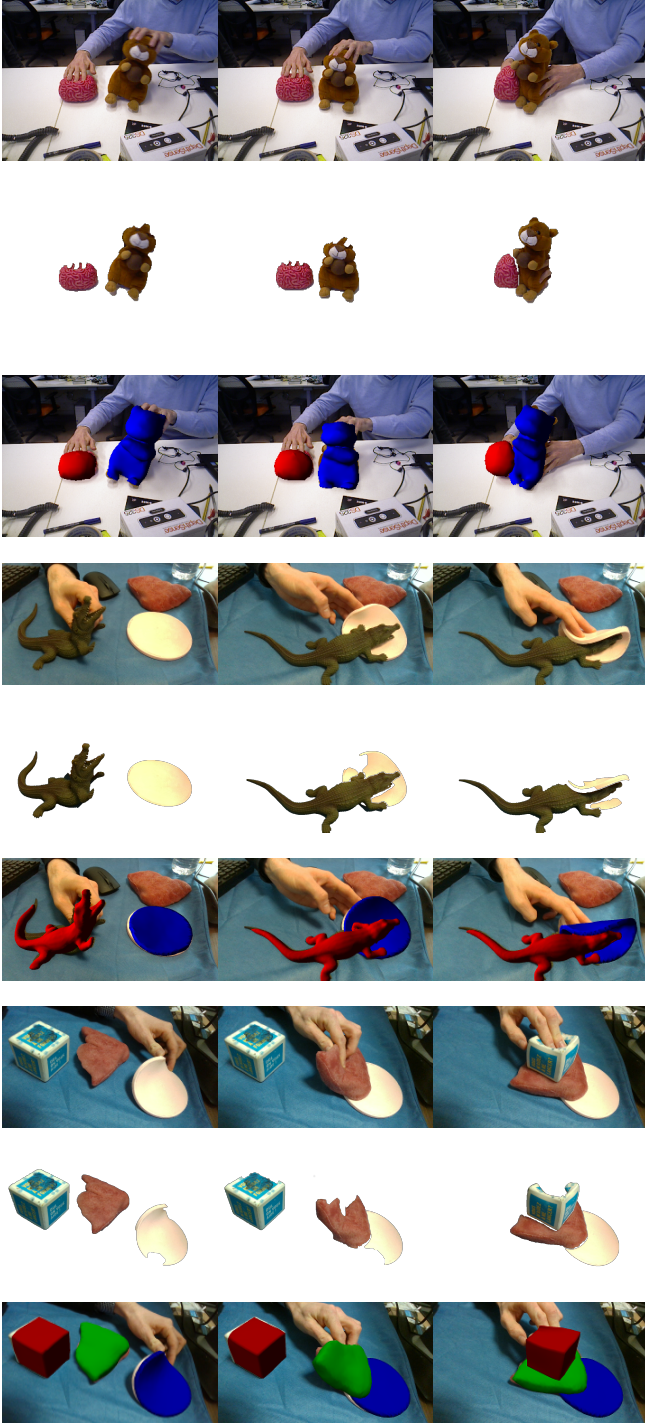


Fig. 3: Results of the proposed registration method for two (first six rows) and three interacting objects (the last three), with the input images, the segmented images (second, fifth and eighth rows), the registered mesh reprojected in the input image. See the provided video for comparisons with rigid registration and collision free approaches.

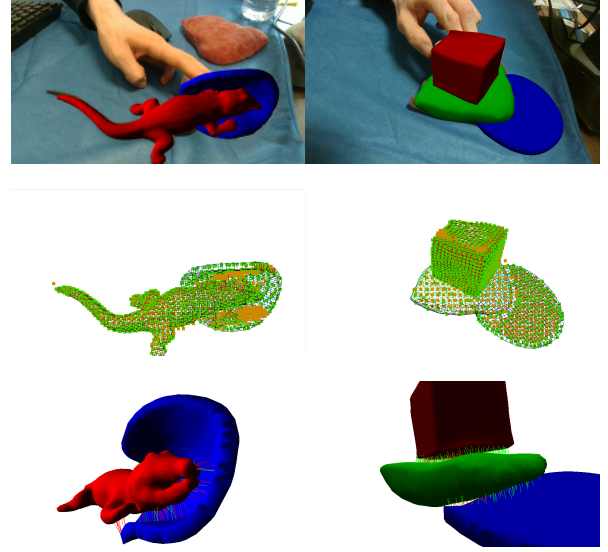


Fig. 4: Visualization of the registered point clouds (brown squares) for the meshes (greens squares as vertices on the visible surfaces) of examples with two and three interacting objects. The displayed segments on the third row show the detected collisions.

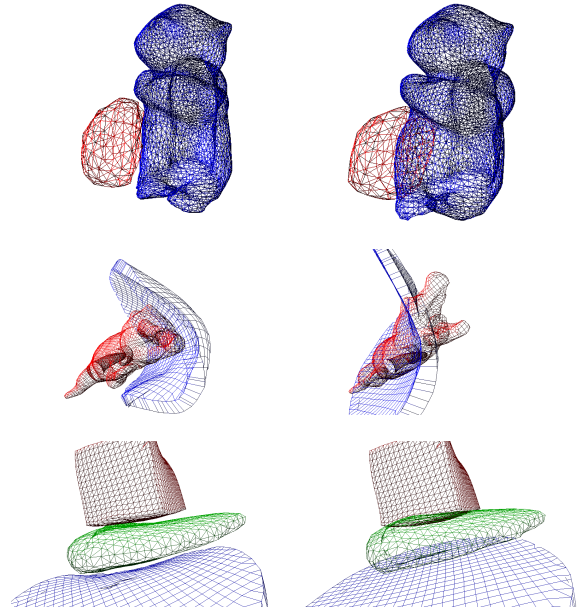


Fig. 5: For the different examples, comparison between registration based on a collision detection model (left column), and without such a model, resulting in large inter-penetrations (right column).



Fig. 6: For the different examples, failure cases when the proposed rigid registration approach is not included in the resolution.

VIII. CONCLUSION

In this paper we propose the first system able to simultaneously register multiple interacting deformable elastic objects. It relies on a co-rotational FEM elastic models to accurately and efficiently model elasticity and on an interaction model consisting in collision detection and response models to cope with contacts between object. Based on RGB-D input data, point cloud matching and registrations processes are achieved on segmented data in a parallel manner, enabling to deal with occlusions. Rigid motions, deformations and contacts of all the considered object are then estimated in a common solver, providing a realistic behaviour and coherently constraining the registration on the involved entities.

The method has been tested on simple scenarios with elastic objects but could be extended to other physical models, to articulated objects, for instance for robotic manipulation applications.

Future works would also aim at proposing a more efficient segmentation process for the different objects, integrating more distinctive point cloud features. Another interesting direction would be to infer realistic contact forces from the registered interaction.

REFERENCES

- [1] J. Allard, F. Faure, H. Courtecuisse, F. Falipou, C. Duriez, and P. G. Kry. Volume contact constraints at arbitrary resolution. In *ACM Transactions on Graphics (TOG)*, volume 29, page 82. ACM, 2010.
- [2] I. Badami, J. Stückler, and S. Behnke. Depth-enhanced hough forests for object-class detection and continuous pose estimation. In *Workshop on Semantic Perception, Mapping and Exploration (SPME)*, 2013.
- [3] A. Bartoli, A. Zisserman, et al. Direct estimation of non-rigid registrations. In *British Machine Vision Conference*, pages 899–908, 2004.
- [4] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992.
- [5] L. D. Cohen and I. Cohen. Deformable models for 3-d medical images using finite elements and balloons. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, pages 592–598. IEEE, 1992.
- [6] R. D. Cook. *Finite element modeling for stress analysis*. Wiley, 1994.
- [7] C. Elbrechter, R. Haschke, and H. Ritter. Bi-manual robotic paper manipulation based on real-time marker tracking and physical modelling. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 1427–1432. IEEE, 2011.
- [8] O. Eitzmuß, M. Keckeisen, and W. Straßer. A fast finite element solution for cloth modelling. In *Computer Graphics and Applications. Proceedings. 11th Pacific Conference on*, pages 244–251. IEEE, 2003.
- [9] F. Faure, C. Duriez, H. Delingette, J. Allard, B. Gilles, S. Marchesseau, H. Talbot, H. Courtecuisse, G. Bousquet, I. Peterlik, et al. Sofa: A multi-model framework for interactive physical simulation. In *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*, pages 283–321. Springer, 2012.
- [10] K. Guo, F. Xu, T. Yu, X. Liu, Q. Dai, and Y. Liu. Real-time geometry, albedo, and motion reconstruction using a single rgb-d camera. *ACM Transactions on Graphics (TOG)*, 36(3):32, 2017.
- [11] H. Hamer, K. Schindler, E. Koller-Meier, and L. Van Gool. Tracking a hand manipulating an object. In *Computer Vision, 2009 IEEE 12th International Conference On*, pages 1475–1482. IEEE, 2009.
- [12] N. Haouchine, S. Cotin, I. Peterlik, J. Dequidt, M. S. Lopez, E. Kerrien, and M.-O. Berger. Impact of soft tissue heterogeneity on augmented reality for liver surgery. *Visualization and Computer Graphics, IEEE Transactions on*, 21(5):584–597, 2015.
- [13] N. Haouchine, J. Dequidt, I. Peterlik, E. Kerrien, M.-O. Berger, and S. Cotin. Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 199–208. IEEE, 2013.
- [14] P. Jiménez, F. Thomas, and C. Torras. 3d collision detection: a survey. *Computers & Graphics*, 25(2):269–285, 2001.
- [15] A. Jordt and R. Koch. Direct model-based tracking of 3d object deformations in depth and color video. *International Journal of Computer Vision*, pages 1–17, 2013.
- [16] O. Kähler, V. A. Prisacariu, and D. W. Murray. Real-time large-scale dense 3d reconstruction with loop closure. In *European Conference on Computer Vision*, pages 500–516. Springer, 2016.
- [17] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.
- [18] K. Kim, V. Lepetit, and W. Woo. Keyframe-based modeling and tracking of multiple 3d objects. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 193–198. IEEE, 2010.
- [19] N. Kyriazis and A. Argyros. Scalable 3d tracking of multiple interacting objects. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3430–3437, 2014.
- [20] A. Malti, R. Hartley, A. Bartoli, and J.-H. Kim. Monocular template-based 3d reconstruction of extensible surfaces with local linear elasticity. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1522–1529. IEEE, 2013.
- [21] T. McInerney and D. Terzopoulos. A finite element model for 3d shape reconstruction and nonrigid motion tracking. In *Computer Vision, 1993. Proceedings., Fourth International Conference on*, pages 518–523. IEEE, 1993.
- [22] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE, 2011.
- [23] I. Oikonomidis, N. Kyriazis, and A. A. Argyros. Full dof tracking of a hand interacting with an object by modeling occlusions and physical constraints. In *2011 International Conference on Computer Vision*, pages 2088–2095. IEEE, 2011.
- [24] I. Oikonomidis, N. Kyriazis, and A. A. Argyros. Tracking the articulated motion of two strongly interacting hands. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1862–1869. IEEE, 2012.
- [25] A. Petit, V. Lippiello, G. A. Fontanelli, and B. Siciliano. Tracking elastic deformable objects with an rgb-d sensor for a pizza chef robot. *Robotics and Autonomous Systems*, 88:187–201, 2017.
- [26] A. Petit, V. Lippiello, and B. Siciliano. Real-time tracking of 3d elastic objects with an rgb-d sensor. *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, 2015.
- [27] T.-H. Pham, A. Kheddar, A. Qammar, and A. A. Argyros. Towards force sensing from vision: Observing hand-object interactions to infer manipulation forces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2810–2819, 2015.
- [28] J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration and realistic augmentation. *Int. Journal of Computer Vision*, 76(2):109–122, Feb. 2007.
- [29] M. Salzmann, J. Pilet, S. Ilic, and P. Fua. Surface deformation models for nonrigid 3d shape recovery. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(8), 2007.
- [30] M. Salzmann and R. Urtasun. Physically-based motion models for 3d tracking: A convex formulation. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2064–2071. IEEE, 2011.
- [31] J. Schulman, A. Lee, J. Ho, and P. Abbeel. Tracking deformable objects with point clouds. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 1130–1137. IEEE, 2013.
- [32] S. Sridhar, F. Mueller, M. Zollhöfer, D. Casas, A. Oulasvirta, and C. Theobalt. Real-time joint tracking of a hand manipulating an object from rgb-d input. In *European Conference on Computer Vision*, pages 294–310. Springer, 2016.
- [33] A. Tejani, D. Tang, R. Kouskouridas, and T.-K. Kim. Latent-class hough forests for 3d object detection and pose estimation. In *European Conference on Computer Vision*, pages 462–477. Springer, 2014.
- [34] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: Recovering 3d shape and nonrigid motion. *Artificial intelligence*, 36(1):91–123, 1988.
- [35] M. Teschner, S. Kimmerle, B. Heidelberger, G. Zachmann, L. Raghupathi, A. Fuhrmann, M.-P. Cani, F. Faure, N. Magnenat-Thalmann, W. Strasser, et al. Collision detection for deformable objects. In *Computer graphics forum*, volume 24, pages 61–81. Wiley Online Library, 2005.
- [36] M. Zollhöfer, M. Nießner, S. Izadi, C. Rehmann, C. Zach, M. Fisher, C. Wu, A. Fitzgibbon, C. Loop, C. Theobalt, et al. Real-time non-rigid reconstruction using an rgb-d camera. *ACM Transactions on Graphics, TOG*, 2014.